

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-227886

(43)Date of publication of application : 15.08.2000

(51)Int.Cl.

G06F 13/00

G06F 12/00

G06F 15/00

G06F 17/30

(21)Application number : 11-028647

(71)Applicant : SHARP CORP

(22)Date of filing : 05.02.1999

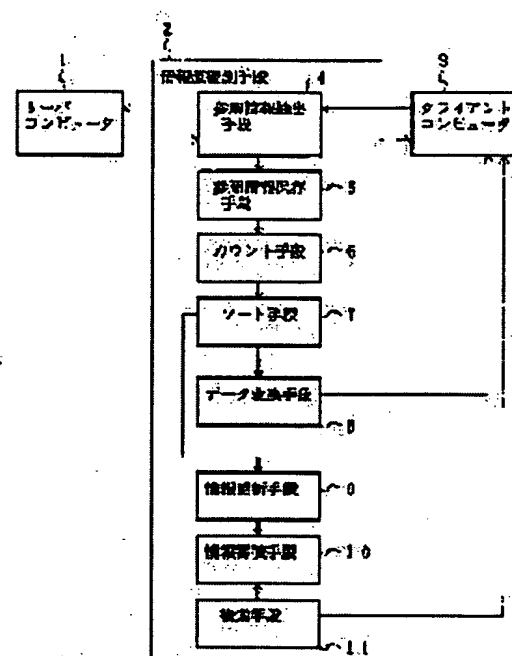
(72)Inventor : DOI KATSUYOSHI

(54) DEVICE AND METHOD FOR INFORMATION SOURCE OBSERVATION AND RECORDING MEDIUM WHERE PROGRAM WHICH EXECUTES INFORMATION SOURCE OBSERVING PROCESSING IS RECORDED

(57)Abstract:

PROBLEM TO BE SOLVED: To sum up what kind of URL pages are referred to by users as reference sources and to present a list of the reference sources to the users in a hypertext system such as WWW.

SOLUTION: A reference information extracting means 4 extracts the URL of a reference source from a request header which is issued by a client computer 3 and data type of a data file from a response header issued by a server computer 1. These pieces of reference information are stored on a reference information storing means 5. When the number of the reference information becomes larger than a specific number, the respective URLs of reference sources are summed up by data type by a counting means 7 and the summed result is presented to the users as the information source list.



LEGAL STATUS

[Date of request for examination]

27.07.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-227886

(P2000-227886A)

(43) 公開日 平成12年8月15日 (2000.8.15)

(51) Int.Cl. ⁷	識別記号	F I	テ-マ-ト ⁷ (参考)
G 0 6 F 13/00	3 5 4	G 0 6 F 13/00	3 5 4 D 5 B 0 7 5
12/00	5 4 6	12/00	5 4 6 A 5 B 0 8 2
15/00	3 1 0	15/00	3 1 0 A 5 B 0 8 5
17/30		15/40	3 1 0 F 5 B 0 8 9
		15/403	3 4 0 B
審査請求 未請求 請求項の数 9 O L (全 12 頁) 最終頁に続く			

(21) 出願番号 特願平11-28647

(22) 出願日 平成11年2月5日 (1999.2.5)

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 土居 克良

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(74) 代理人 100080034

弁理士 原 謙三

Fターム(参考) 5B075 ND04 ND36 PQ02 PQ46 PQ72

PR03

5B082 FA11 FA12 FA13

5B085 AC14 BG07

5B089 GA12 GA23 GB03 HA10 JA22

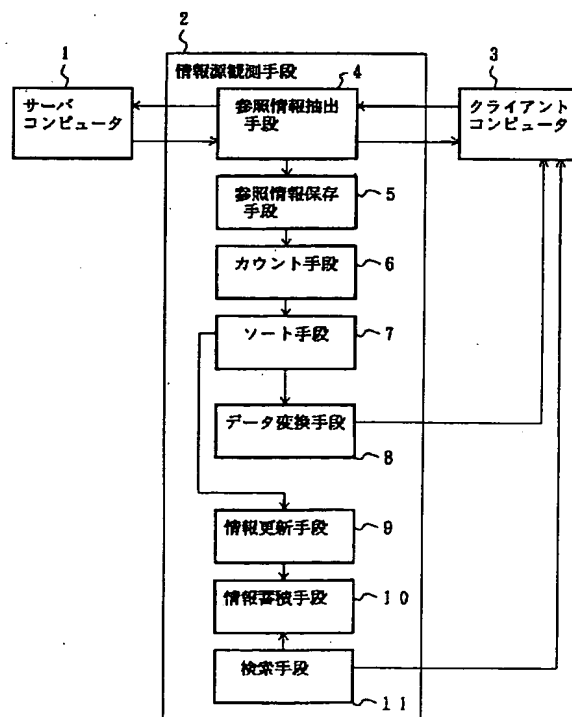
JA37 KA14 KB07 KC30

(54) 【発明の名称】 情報源観測装置および情報源観測方法ならびに情報源観測処理を実行させるプログラムを記録した記録媒体

(57) 【要約】

【課題】 WWWなどのハイパーテキストシステムにおいて、ユーザがどのようなページを参照元としてURLページを参照しているかを集計し、その参照元のリストをユーザに提示する。

【解決手段】 参照情報抽出手段4によって、クライアントコンピュータ3で発行される要求ヘッダから参照元のURLを、サーバコンピュータ1で発行される応答ヘッダからデータファイルのデータタイプを抽出し、これらの参照情報を参照情報保存手段5に保存する。そして、参照情報が一定数以上となった場合に、カウント手段6およびソート手段7によって、データタイプ毎に、各参照元のURLを集計し、集計結果を情報源リストとしてユーザに提示する。



【特許請求の範囲】

【請求項1】ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際に用いられる情報源観測装置において、

クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する参照情報保存手段と、

上記参照情報を、参照元となる各URL毎に集計するカウント手段とを有していることを特徴とする情報源観測装置。

【請求項2】クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われる際に、クライアントコンピュータから発行される要求ヘッダから参照元となるURLの情報を抽出する参照情報抽出手段をさらに備えていることを特徴とする請求項1記載の情報源観測装置。

【請求項3】上記参照情報抽出手段が、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われた際に、サーバコンピュータから発行される応答ヘッダから、上記データファイルのデータタイプを抽出することを特徴とする請求項2記載の情報源観測装置。

【請求項4】上記参照情報を、データタイプ別に、参照元となる各URL毎に集計するソート手段をさらに備えていることを特徴とする請求項3記載の情報源観測装置。

【請求項5】上記参照情報保存手段は、クライアントコンピュータからアクセスがあったデータファイルに付随するテキストをさらに保存することを特徴とする請求項1記載の情報源観測装置。

【請求項6】クライアントコンピュータからアクセスがあったデータファイルをキャッシュとして所定量蓄積する情報蓄積手段と、

上記参照情報の集計結果に基づいて、上記キャッシュのうち、参照元となるURLのデータファイル、および、参照元となるURLのデータファイルに所定のリンクレベルでリンクされているデータファイルを、所定の時間毎に、該当するサーバコンピュータにアクセスすることによって更新する情報更新手段とをさらに備えていることを特徴とする請求項1または4記載の情報源観測装置。

【請求項7】上記キャッシュ内のデータファイルに対して、検索処理を行う検索手段をさらに備えていることを特徴とする請求項6記載の情報源観測装置。

【請求項8】ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測方法において、

ファイルの参照元のURLを参照情報として保存するステップと、

上記参照情報を、参照元となる各URL毎に集計するステップとを有していることを特徴とする情報源観測方法。

【請求項9】ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測処理を実行させるプログラムを記録した記録媒体において、

クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する処理と、

上記参照情報を、参照元となる各URL毎に集計する処理とをコンピュータに実行させるためのプログラムを記録した、コンピュータ読み取り可能であることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、例えばサーバコンピュータとクライアントコンピュータとがネットワークで結ばれたシステムにおいて、サーバコンピュータ上にあるハイパーリンク構造をもったデータファイルをクライアントコンピュータ上において閲覧する際に利用される情報源観測装置および情報源観測方法ならびに情報源観測処理を実行させるプログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】昨今、複数のサーバコンピュータおよび複数のクライアントコンピュータがネットワークで結ばれ、各サーバコンピュータにハイパーテキスト構造のマルチメディアデータが記憶されており、各クライアントコンピュータにおいて、ブラウザと呼ばれるソフトウェアによって、このようなマルチメディアデータを閲覧することが可能なシステムが広く普及している。このようなシステムの例としては、例えば、インターネットにおけるWorld Wide Web (WWW) と呼ばれるシステムなどが挙げられる。

【0003】マルチメディアデータを含む文書は、例えばHyper Text Markup Language (HTML) と呼ばれる記述言語によって記述されており、テキスト文書、静止画、動画、音楽データ、およびJava (登録商標) アプレットなどのアプリケーションプログラムなどを含むことが可能になっている。このような文書(以下、HTMLページと称する)およびマルチメディアデータには、それぞれURL (Uniform Resource Locator) と呼ばれる固有のアドレスが割り当てられている。ユーザは、ブラウザ上においてURLを指定することによって、所望のHTMLページあるいはマルチメディアデータにアクセスすることができる。また、HTMLページには、他のペ

ージやマルチメディアデータへアクセスするためのリンクなども埋め込まれており、ユーザは、このリンクをブラウザ上でポイントすることによって、リンク先のページやマルチメディアデータに移動することも可能となっている。

【0004】クライアントコンピュータにおけるHTML表示ソフトウェアであるブラウザとしては、例えば、Netscape Communications 社のNetscape Communicator（登録商標）や、Microsoft 社のInternet Explorer（登録商標）などの製品が広く普及している。これらのブラウザでは、アクセスしたHTMLページおよびマルチメディアデータのURL、アクセス日時、およびタイトルなどが、履歴データとしてクライアントコンピュータ内のハードディスクに記憶することが可能になっている。そして、ユーザは、過去にアクセスしたHTMLページあるいはマルチメディアデータに再びアクセスしたい時に、この履歴データを参照することによって、容易に所望のHTMLページあるいはマルチメディアデータにアクセスすることが可能となる。

【0005】また、例えばNetscape Communicator では、履歴データに基づいて、過去にアクセスしたHTMLページおよびマルチメディアデータのURLを、頻度順、あるいは日時順などに並べ変えて表示することが可能となっている。

【0006】さらに、例えば特開平10-143519号公報には、ユーザが過去にアクセスしたURLに対して、頻度や視聴時間をもとに順序づけを行い、その結果を表示する方法および装置が開示されている。

【0007】また、例えば特開平9-204347号公報、特開平10-21134号公報には、サーバコンピュータとクライアントコンピュータとの間でURLの中継を行うゲートウェイコンピュータにおいて、ゲートウェイコンピュータに中継キャッシュが内蔵されている場合、過去に中継したURLに対して、その頻度の算出を行ってリストを作成するとともに、頻度順にゲートウェイコンピュータが自発的にキャッシュの更新を行う方法が開示されている。

【0008】これらの方式に共通する点としては、過去にアクセスしたHTMLページおよびマルチメディアデータのURLに関して、その頻度に注目し、それらを統計処理することによって頻度を算出し、頻度の高いHTMLページおよびマルチメディアデータは、ユーザの要

求度が高いと判断している点である。

【0009】

【発明が解決しようとする課題】ここで、情報を提供する複数のサーバコンピュータと、ゲートウェイコンピュータと、クライアントコンピュータとからなるシステムを想定する。なお、ゲートウェイコンピュータとは、異なるネットワーク／システムを相互接続するためのコンピュータを示すものである。

【0010】そして、例えば、図8に示すように、あるサーバコンピュータ上に、URLが“http://www.news/”で示されるHTMLページが存在し、別のサーバコンピュータ上に、URLが“http://www.hello.nara/”で示されるHTMLページが存在するとする。

【0011】URLが“http://www.news/”で示されるHTMLページは、あるニュース情報（「日々新聞」）を提供するページであり、「新刊情報（“http://www.news/1.html”）」、「天気予報（“http://www.news/2.html”）」、「A社新製品（“http://www.news/3.html”）」、「B社新製品（“http://www.news/4.html”）」、および「C社新製品（“http://www.news/5.html”）」の5つのHTMLページへのリンクがはられてある。

【0012】また、URLが“http://www.hello.nara/”で示されるHTMLページは、奈良県地域情報を提供するページであり、「株式情報（“http://www.a.mov”）」および「道路情報（“http://www/b.mov”）」の2つの動画データ、および「お知らせ（“http://www/index.html”）」のHTMLページにリンクがはられてある。

【0013】ここで、あるユーザが、クライアントコンピュータにおいて、まず、URLが“http://www.news/”で示されるHTMLページにアクセスし、このページからリンクされている5つのHTMLページ（テキストデータ）を閲覧し、その後、URLが“http://www.hello.nara/”で示されるHTMLページにアクセスし、このページからリンクされている2つの動画データ、および1つのHTMLページ（テキストデータ）を閲覧したとする。

【0014】このとき、以下に示す表1のようなアクセスログが、ゲートウェイコンピュータに記録される。

【0015】

【表1】

URL	応答コード	Content-type	タイトル情報
http://www.news/	200	text/html	日々新聞
http://www.news/1.html	200	text/html	新刊情報
http://www.news/2.html	200	text/html	天気予報
http://www.news/3.html	200	text/html	A社新製品情報
http://www.news/4.html	200	text/html	B社新製品情報
http://www.news/5.html	200	text/html	C社新製品情報
http://www.hello.nara/	200	text/html	奈良県地域情報
http://www/a.mov	200	movie	株式情報
http://www/b.mov	200	movie	道路情報

【0016】なお、表1において、URLが“http://www.news/”で示されるHTMLページが一回しかカウントされていないのは、次のような理由によるものである。例えば、URLが“http://www.news/”で示されるHTMLページから、リンク先である「新刊情報（“http://www.news/1.html”）」のページへ移動し、閲覧後、ブラウザのバックボタン（戻るボタン）を押すことによって再び“http://www.news/”のページに戻り、同様にして他の4つのリンク先に移動したとする。この際に、ブラウザには、通常、キャッシュが一定量蓄積されており、「新刊情報（“http://www.news/1.html”）」のページから、“http://www.news/”のページに戻る際には、このキャッシュから“http://www.news/”のページのデータが取り出されることになる。この場合、ゲートウェイコンピュータには、URLの取得要求が伝達されないで、ゲートウェイコンピュータのアクセスログには、URLが“http://www.news/”で示されるHTMLページが、一回しか現れてこないことになる。

【0017】表1に示す集計結果によれば、どのURLも1回の出現頻度であるから、上記のように、出現頻度に基づいてユーザの要求度を判断する方式によれば、どのURLもユーザの要求度は同等とみなされることになる。しかしながら、実際には、ユーザは、後日、URLが“http://www.news/”で示されるHTMLページや、URLが“http://www.hello.nara/”で示されるHTMLページに再びアクセスし、新しいリンクが作られていれば、その新しいリンクにアクセスする、というアクセス動作を行う可能性が高いと予想される。この際に、すでにアクセス済みのURLである“http://www.news/1.html”や、“http://www.news/2.html”などへ再びアクセスすることはほとんどないものと予想される。

【0018】すなわち、上記のように、出現頻度に基づいてユーザの要求度を判断する方式の場合には、実際のユーザの要求度を反映した判定を行っていないことになる。よって、このような方式によって示された結果には、無駄なデータが多く含まれることになり、ユーザにとって有用な情報が示されているとはいえない場合が

ありうる。

【0019】本発明は上記の問題点を解決するためになされたもので、その目的は、WWWなどのハイパーテキストシステムにおいて、ユーザがどのようなページを参照元としてURLページを参照しているかを集計し、その参照元のリストをユーザに提示することにある。

【0020】

【課題を解決するための手段】上記の課題を解決するために、請求項1記載の情報源観測装置は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際に用いられる情報源観測装置において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する参照情報保存手段と、上記参照情報を、参照元となる各URL毎に集計するカウント手段とを有していることを特徴としている。

【0021】上記の構成によれば、参照情報保存手段によって、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存し、カウント手段によって、上記参照情報を、参照元となる各URL毎に集計するので、参照元として頻繁に利用されるURLを把握することが可能となる。すなわち、ユーザにとって有用なURLを的確に提示することが可能となる。

【0022】請求項2記載の情報源観測装置は、請求項1記載の構成において、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われる際に、クライアントコンピュータから発行される要求ヘッダから参照元となるURLの情報を抽出する参照情報抽出手段をさらに備えていることを特徴としている。

【0023】上記の構成によれば、参照情報抽出手段によって、クライアントコンピュータから発行される要求ヘッダから参照元となるURLの情報を抽出しており、このような要求ヘッダは、HTTP1.0の仕様に基づいているものであるため、特別にクライアントコンピュ

ータにデータを要求することなく、参照元となるURLの情報を取得することができる。

【0024】請求項3記載の情報源観測装置は、請求項2記載の構成において、上記参照情報抽出手段が、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われた際に、サーバコンピュータから発行される応答ヘッダから、上記データファイルのデータタイプを抽出することを特徴としている。

【0025】上記の構成によれば、参照情報抽出手段が、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われた際に、サーバコンピュータから発行される応答ヘッダから、上記データファイルのデータタイプを抽出しており、このような応答ヘッダは、HTTP 1.0の仕様に基づいているものであるため、特別にサーバコンピュータにデータを要求することなく、データファイルのデータタイプを取得することができる。

【0026】請求項4記載の情報源観測装置は、請求項3記載の構成において、上記参照情報を、データタイプ別に、参照元となる各URL毎に集計するソート手段をさらに備えていることを特徴としている。

【0027】上記の構成によれば、ソート手段によって、参照情報を、データタイプ別に、参照元となる各URL毎に集計するので、データタイプ毎に、参照元として頻繁に利用されるURLを把握することが可能となる。すなわち、データタイプに応じて、ユーザにとって有用なURLを的確に提示することが可能となる。

【0028】請求項5記載の情報源観測装置は、請求項1記載の構成において、上記参照情報保存手段が、クライアントコンピュータからアクセスがあったデータファイルに付随するテキストをさらに保存することを特徴としている。

【0029】上記の構成によれば、参照情報保存手段によって、クライアントコンピュータからアクセスがあったデータファイルに付随するテキストが参照情報としてさらに保存されるので、ユーザにとって有用なURLを提示する際に、ユーザの嗜好する内容をも提示することができる。

【0030】請求項6記載の情報源観測装置は、請求項1または4記載の構成において、クライアントコンピュータからアクセスがあったデータファイルをキャッシュとして所定量蓄積する情報蓄積手段と、上記参照情報の集計結果に基づいて、上記キャッシュのうち、参照元となるURLのデータファイル、および、参照元となるURLのデータファイルに所定のリンクレベルでリンクされているデータファイルを、所定の時間毎に、該当するサーバコンピュータにアクセスすることによって更新する情報更新手段とをさらに備えていることを特徴としている。

【0031】上記の構成によれば、情報蓄積手段によって、クライアントコンピュータからアクセスがあったデータファイルをキャッシュとして所定量蓄積し、情報更新手段によって、参照情報の集計結果に基づいて、上記キャッシュのうち、参照元となるURLのデータファイル、および、参照元となるURLのデータファイルに所定のリンクレベルでリンクされているデータファイルを、所定の時間毎に、該当するサーバコンピュータにアクセスすることによって更新するので、ユーザが参照元として数多く利用するURLを起点にした情報収集となり、ユーザが利用する可能性の高い、的確な情報収集が可能となる。これにより、参照元のURLからリンクされているデータに対してユーザがアクセスした場合にも、キャッシュデータの中からデータを取り出すことが可能となり、アクセススピードの向上およびネットワークトラフィックの低減を図ることができる。

【0032】請求項7記載の情報源観測装置は、請求項6記載の構成において、上記キャッシュ内のデータファイルに対して、検索処理を行う検索手段をさらに備えていることを特徴としている。

【0033】上記の構成によれば、検索手段によって、上記キャッシュ内のデータファイルに対して、検索処理を行うので、ユーザの情報源をもとに情報収集されたファイルオブジェクトに対して検索が行われることになり、一般のインターネット検索エンジンなどに比べて、情報の鮮度が高く、かつ、ユーザのニーズに適した検索結果を提供することができる。

【0034】請求項8記載の情報源観測方法は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測方法において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存するステップと、上記参照情報を、参照元となる各URL毎に集計するステップとを有していることを特徴としている。

【0035】上記の方法によれば、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存し、上記参照情報を、参照元となる各URL毎に集計するので、参照元として頻繁に利用されるURLを把握することが可能となる。すなわち、ユーザにとって有用なURLを的確に提示することが可能となる。

【0036】請求項9記載の情報源観測処理を実行させるプログラムを記録した記録媒体は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測処理を実行させるプログラムを記録した記録媒体において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する処理と、上記参照情報を、参照

元となる各URL毎に集計する処理とをコンピュータに実行させるためのプログラムを記録した、コンピュータ読み取り可能であることを特徴としている。

【0037】上記の構成によれば、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存し、上記参照情報を、参照元となる各URL毎に集計するプログラムが記録されているので、参照元として頻繁に利用されるURLを把握することが可能なプログラムを提供することができる。すなわち、ユーザにとって有用なURLを的確に提示することが可能なプログラムを提供することができる。

【0038】

【発明の実施の形態】本発明の実施の一形態について図1ないし図7に基づいて説明すれば、以下のとおりである。

【0039】図1は、本実施形態に係るコンピュータネットワークシステムの概略構成を示すブロック図である。該コンピュータネットワークシステムは、サーバコンピュータ1、情報源観測手段（情報源観測装置）2、およびクライアントコンピュータ3とを備えている。なお、図中において、サーバコンピュータ1は1つしか記載されていないが、サーバコンピュータ1としては、インターネット上に存在する無数のサーバコンピュータが該当することになる。

【0040】サーバコンピュータ1は、HTMLページや各種マルチメディアデータを保持しており、例えばHTTP (Hyper Text Transfer Protocol) デモンと呼ばれるサーバソフトウェアによって、これらのデータの管理、および外部コンピュータからのアクセスの管理を行っている。

【0041】クライアントコンピュータ3は、HTMLページや各種マルチメディアデータを表示可能なブラウザとよばれるソフトウェアを備えており、ユーザはこのクライアントコンピュータ3上においてブラウザを操作することによって、所望のHTMLページや各種マルチメディアデータを閲覧することができる。

【0042】情報源観測手段2としては、例えばファイアーウォール上に設けられるゲートウェイコンピュータや、インターネット上に設けられるHTTPプロキシサーバコンピュータなどが該当し、HTTPリクエストと応答の中継を行うことを基本としている。この情報源観測手段2は、参照情報抽出手段4、参照情報保存手段5、カウント手段6、ソート手段7、データ変換手段8、情報更新手段9、情報蓄積手段10、および検索手段11を備えている。これらの各手段についての説明は後述する。

【0043】また、図2に示すように、情報源観測手段2を、クライアントコンピュータ3に内蔵させた構成とすることも可能である。このように、クライアントコンピュータ3に内蔵可能なHTTPプロキシサーバとして

は、シャープ株式会社製のシャープモバイルプロキシ（商品名）などがある。

【0044】さらに、情報源観測手段2を、クライアントコンピュータ3におけるブラウザの一部として、ソフトウェア的に構成することも可能である。

【0045】図3は、クライアントコンピュータ3の概略構成を示すブロック図である。図3に示すように、クライアントコンピュータ3は、CPU (Central Processing Unit) 12、RAM (Random Access Memory) などで構成されるメモリ13、ハードディスクやフラッシュメモリなどで構成される不揮発性メモリ14、外部ネットワークとのインターフェースとなるネットワークI/O (Input/Output) 15、キーボードやマウスなどで構成される入力装置16、および表示装置17を備えており、これらがバス18によって接続されている。このような構成のクライアントコンピュータ3は、一般的にパーソナルコンピュータと呼ばれるコンピュータによって構成されることになる。

【0046】ここで、クライアントコンピュータ3におけるブラウザによってWWW上のHTMLページを閲覧する際の、情報源観測手段2の動作について説明する。

【0047】まず、あるHTMLページをブラウザ上に表示させている状態から、そのHTMLページに埋め込まれているリンクをマウスカースルでポイントし、クリックすることによって、リンク先のURLにアクセスした場合を想定する。例えば、図8に示す例において、URLが“http://www.news/”で示されるHTMLページを表示させている状態から、「新刊情報」のHTMLページを示すリンクをクリックしたとする。この時、クライアントコンピュータ3において、ネットワークI/O 15を介して、情報源観測手段2に対してTCP/IP (Transmission Control Protocol/Internet Protocol) のコネクションがオープンされ、図4に示すような要求ヘッダが発行される。

【0048】この要求ヘッダに含まれるReferer ヘッダは、ブラウザによって発行されているものであり、要求されているURLの参照元情報が示されている。図4に示す例では、現在要求しているURL “http://www.news/1.html”が、URLが“http://www.news/”のHTMLページからのリンクによって要求されていることを示している。このような要求ヘッダは、HTTP 1.0の仕様で規定されている標準ヘッダである。このような要求ヘッダの情報は、通常サーバコンピュータにて記録されることを想定して定義されているものである。

【0049】情報源観測手段2は、上記のような要求ヘッダを受けて、URLが“www.news”で示されるサーバコンピュータ1にコネクションを張り、該当するファイル (1.html) を受信し、クライアントコンピュータ3のブラウザに転送する。この際に、情報源観測手段2がサーバコンピュータ1から受信する応答データは、図5に

示すように、応答ヘッダとテキストデータ（HTMLデータ）とからなっている。

【0050】図5において、応答ヘッダにあるContent-Typeヘッダは、後に続くデータのマルチメディアタイプを示しており、図5に示す例では、後に続くデータがテキストデータ（HTMLデータ）であることを示している。

【0051】本実施形態における情報源観測手段2は、上記のような要求ヘッダおよび応答ヘッダに含まれる情報を利用して、アクセスしたURLのうちで正常応答を示したURLの参照元情報およびマルチメディアタイプを取得し、アクセスログの項目として記録している。これにより、例えば、アクセスログにおいて参照元情報に注目して集計を行えば、アクセスしたURLの参照元となったURLを頻度順に並びかえて、情報源リストとしてユーザに提示することが可能となる。

【0052】また、参照元情報は、マルチメディアデータタイプ毎に集計することも可能となる。例えば、テキストデータ、動画データ、音楽データなどのデータタイプ毎に分別して、情報源リストとしてユーザに提示することも可能となる。

【0053】また、情報源観測手段2として、HTTPプロキシサーバコンピュータやゲートウェイコンピュータなどのような、キャッシュを蓄積するタイプのコンピュータの場合、情報源リストに基づいて、参照元として頻度の高いURLに対して、キャッシュを自動的に更新する構成とすることも可能である。

【0054】次に、情報源観測手段2における処理の流れを、図2を参照しながら、図6に示すフローチャートに基づいて説明する。なお、以下の説明において、情報源観測手段2は、クライアントコンピュータ3に内蔵されているものとしているが、クライアントコンピュータ3の外部に設けられた構成においても、ほぼ同様の処理

が行われる。

【0055】まず、情報源観測手段2は、ネットワークI/O15を用いてサーバソケットをオープンし、中継要求を受け付ける待機状態に入る（ステップ1、以降、S1のように表記する）。中継要求を受信した場合には、S2に進む。なお、上記のように、情報源観測手段2がクライアントコンピュータ3に内蔵されている場合には、情報源観測手段2は、メモリ13を経由してブラウザ3Aと通信する。

【0056】次に、情報源観測手段2は、中継要求を受け付けると、参照情報抽出手段4によって、要求ヘッダから、参照元情報であるRefererヘッダを抽出し、この内容をメモリ13に一時記憶させる（S2）。

【0057】次に、情報源観測手段2は、サーバコンピュータ1に対してコネクションを開いて要求ヘッダを中継する（S3）。この際に、要求したURLが存在しない場合には、エラーメッセージがクライアントコンピュータ3の表示装置17によってユーザに提示され、メモリ13に一次記憶させたRefererヘッダの情報が破棄される（S4）。

【0058】要求したURLが存在する場合（応答コードが正常である場合）には、そのURLに該当するサーバコンピュータ1からの応答データを受け取り、クライアントコンピュータ3におけるブラウザ3Aに該応答データを送信する。また、これと同時に、参照情報保存手段5によって、要求したURLが、応答ヘッダに含まれるContent-typeヘッダに基づいて検出されるデータタイプ、Refererヘッダに基づいて検出される参照元情報、および参照元のタイトルとともに、参照情報として不揮発性メモリ14に記録され、保存される（S5）。この保存された参照情報の保存データの例を表2に示す。

【0059】

【表2】

URL	出現 頻度	データ タイプ	参照元情報 (Referer)	参照元 タイトル
http://www.news/	1			
http://www.news/1.html	1	text/html	http://www.news/	日々新聞
http://www.news/2.html	1	text/html	http://www.news/	日々新聞
http://www.news/3.html	1	text/html	http://www.news/	日々新聞
http://www.news/4.html	1	text/html	http://www.news/	日々新聞
http://www.news/5.html	1	text/html	http://www.news/	日々新聞
http://www.hello.nara/	1			
http://www/a.mov	1	movie	http://www.hello/nara/	奈良県地域情報
http://www/b.mov	1	movie	http://www.hello/nara/	奈良県地域情報
http://www/index.html	1	text/html	http://www.hello/nara/	奈良県地域情報

【0060】ここで、保存されるデータのデータタイプとしてはtext/htmlなどがあるが、これ以外にあらかじめ指定されたデータタイプだけを保存してもよい。なお、Content-typeヘッダがtext/htmlであるときにおいて、応答データのHTMLデータを構文解析し、<title>; タグで囲まれた部分を抽出することによって、該当ページのタイトルとして同時に保存してもよい。

【0061】上記のような参照情報の保存処理は、保存データが一定数になるまで繰り返して行われる（S6）。そして、保存データが一定数に達したら、次のステップ（S7）に進む。

【0062】一定数に達した保存データに対して、カウント手段6によって、参照元情報毎に参照に用いられた回数を集計し、集計データを作成する（S7）。そして、ソート手段7によって上記集計データをデータタイプ別にソートする（S8）。このソート結果の例を表3および表4に示す。なお、表3は、表2に示す参照情報の保存データの例において、データタイプとしてtext/htmlまたはmovieを集計した結果を示し、表4は、データタイプとしてtext/htmlのみを集計した結果を示している。

【0063】

【表3】

データタイプとしてtext/html またはmovie を集計した場合

URL	参照頻度
http://www.news/	5
http://www.hello.nara/	3

【0064】

【表4】

データタイプとしてtext/html のみを集計した場合

URL	参照頻度	タイトル
http://www.news/	5	日々新聞
http://www.hello.nara/	1	奈良県地域情報

【0065】このようにソートされた参照情報のデータを、データ変換手段によって、例えばHTMLデータなどの、ユーザがブラウザで表示することが可能なデータ形式に変換し、情報源リストとして表示装置17に表示することによって、ユーザに提示する（S9）。このような情報源リストの例を図7に示す。

【0066】情報源リストには、参照元情報のページのタイトルを示す情報が付随しているので、クライアントコンピュータ3を使用するユーザが、どのような情報を求めることを習慣としているかを把握することができる。すなわち、この情報源リストは、嗜好データベースとして利用できる。例えば、図7に示す例では、ユーザが日々新聞、奈良県の地域情報などを情報ソースとして

重要と考えていることがわかる。また、各データタイプ毎に参照元情報の集計が提示されているので、データタイプ毎に、重要となる参照元情報を把握することができる。

【0067】ここで、情報源観測手段2が、アクセスしたURLのデータをキャッシュとして情報貯蓄手段10に一定量蓄積するものである場合を考える。この場合、情報更新手段9が、所定の時間毎に、上記のソートされた参照情報のデータに基づいて、サーバコンピュータ1から情報を自発的に取得し、情報蓄積手段10に蓄積されたデータの更新が行われる（S10）。このような情報更新処理は、ソートされた参照情報のデータにおいて、例えば全てのデータタイプにおける参照元として頻度の高い順、あるいは、ユーザが指定したデータタイプにおける参照元として頻度の高い順などに基づいて行われる。

【0068】この情報更新手段9による情報更新処理において、情報源リストに掲載されたURL群を対象にリンクレベル2以上でファイルオブジェクトの自動収集する処理を行えば、参照元としてのURLからリンクされているデータ、すなわち、ユーザがアクセスする可能性の高いデータをも自動的に情報蓄積手段10に蓄積されることになる。すなわち、リンクレベルを深くすれば、参照元としてのURLからリンクされている広い範囲の関連情報まで自動収集することになるが、ユーザが参照元として数多く利用するURLを起点にした情報収集となるので、ユーザが利用する可能性の高い、的確な情報収集が可能となる。これにより、参照元のURLからリンクされているデータに対してユーザがアクセスした場合にも、キャッシュデータの中からデータを取り出すことが可能となり、アクセススピードの向上およびネットワークトラフィックの低減を図ることができる。

【0069】さらに、上記のように、情報源リストに掲載されたURL群を対象に所定のリンクレベル以上で自動収集したファイルオブジェクトに対して、検索手段11により検索をかけ、ユーザに結果を提示することも可能である。この場合、ユーザの情報源をもとに情報収集されたファイルオブジェクトに対して検索が行われるので、一般のインターネット検索エンジンなどに比べて、情報の鮮度が高く、かつ、ユーザのニーズに適した検索結果を提供することができる。

【0070】以上のようなS1からS10のステップを、情報源観測手段2は繰り返して行うことになる。

【0071】このように、参照情報を集計し、出現頻度順にソートした結果において、ユーザが参照元として頻繁に利用したURLは上位にランクされることになる。したがって、このようなソート結果をユーザに提示すれば、ユーザは、所望の情報をどこから得ることが出来るかを、自分の情報源リストから発見することが可能となり、この情報源リストからリンクをたどることによ

て、容易に情報を得ることが可能となる。

【0072】ここで、情報源観測手段2がクライアントコンピュータ3に内蔵されている構成の場合において、情報源観測手段2が備える各手段が、クライアントコンピュータ2内でどのように構成されるかについて、その一例を説明する。まず、上記のS1からS10までの処理を行うプログラムコード、およびこのプログラムコードを実行可能なOS (Operating System) コードが、ファイル形態で不揮発性メモリ14に格納されている。また、CPUアドレス空間にマッピング可能な、ROM (Read Only Memory) などの不揮発性メモリ14に、電源がONされた時に、上記のOSコードをメモリ13にロードさせる処理などを行うコードが格納されている。すなわち、情報源観測手段2が備える各手段は、不揮発性メモリに格納されたプログラムコードおよびOSコードが、メモリ12にロードされて、CPU12によって実行されることによって実現される。また、情報源観測手段2が、クライアントコンピュータ3とは別の、外部のコンピュータに設けられている場合にも、上記とほぼ同様の構成によって実現されることになる。

【0073】また、情報源観測手段2が、例えばインターネット対応テレビやインターネット対応電話などの情報家電機器に内蔵されている場合には、例えば、上記のような処理を行うプログラムコードが直接ROMなどの不揮発性メモリに記憶され、CPUがこのプログラムコードをROMから読み出すことによって実行されることになる。また、このような情報家電機器においても、OSを備えた構成とし、OS上でプログラムを実行する構成とすることも可能である。

【0074】

【発明の効果】以上のように、請求項1の発明に係る情報源観測装置は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際に用いられる情報源観測装置において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する参照情報保存手段と、上記参照情報を、参照元となる各URL毎に集計するカウント手段とを有している構成である。

【0075】これにより、参照元として頻繁に利用されるURLを把握することが可能となるという効果を奏する。すなわち、ユーザにとって有用なURLを的確に提示することが可能となるという効果を奏する。

【0076】請求項2の発明に係る情報源観測装置は、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われる際に、クライアントコンピュータから発行される要求ヘッダから参照元となるURLの情報を抽出する参照情報抽出手段をさらに備えている構成である。

【0077】これにより、請求項1の構成による効果に

加えて、特別にクライアントコンピュータにデータを要求することなく、参照元となるURLの情報を取得することができるという効果を奏する。

【0078】請求項3の発明に係る情報源観測装置は、上記参照情報抽出手段が、クライアントコンピュータにおいて、サーバコンピュータ上のデータファイルにアクセスが行われた際に、サーバコンピュータから発行される応答ヘッダから、上記データファイルのデータタイプを抽出する構成である。

【0079】これにより、請求項2の構成による効果に加えて、特別にサーバコンピュータにデータを要求することなく、データファイルのデータタイプを取得することができるという効果を奏する。

【0080】請求項4の発明に係る情報源観測装置は、上記参照情報を、データタイプ別に、参照元となる各URL毎に集計するソート手段をさらに備えている構成である。

【0081】これにより、請求項3の構成による効果に加えて、データタイプ毎に、参照元として頻繁に利用されるURLを把握することが可能となるという効果を奏する。すなわち、データタイプに応じて、ユーザにとって有用なURLを的確に提示することが可能となるという効果を奏する。

【0082】請求項5の発明に係る情報源観測装置は、上記参照情報保存手段が、クライアントコンピュータからアクセスがあったデータファイルに付随するテキストをさらに保存する構成である。

【0083】これにより、請求項1の構成による効果に加えて、ユーザにとって有用なURLを提示する際に、ユーザの嗜好する内容をも提示することができるという効果を奏する。

【0084】請求項6の発明に係る情報源観測装置は、クライアントコンピュータからアクセスがあったデータファイルをキャッシュとして所定量蓄積する情報蓄積手段と、上記参照情報の集計結果に基づいて、上記キャッシュのうち、参照元となるURLのデータファイル、および、参照元となるURLのデータファイルに所定のリンクレベルでリンクされているデータファイルを、所定の時間毎に、該当するサーバコンピュータにアクセスすることによって更新する情報更新手段とをさらに備えている構成である。

【0085】これにより、請求項1または4の構成による効果に加えて、ユーザが参照元として数多く利用するURLを起点にした情報収集となり、ユーザが利用する可能性の高い、的確な情報収集が可能となる。これにより、参照元のURLからリンクされているデータに対してユーザがアクセスした場合にも、キャッシュデータの中からデータを取り出すことが可能となり、アクセススピードの向上およびネットワークトラフィックの低減を図ることができるという効果を奏する。

【0086】請求項7の発明に係る情報源観測装置は、上記キャッシュ内のデータファイルに対して、検索処理を行う検索手段をさらに備えている構成である。

【0087】これにより、請求項6の構成による効果に加えて、ユーザの情報源をもとに情報収集されたファイルオブジェクトに対して検索が行われることになり、一般のインターネット検索エンジンなどに比べて、情報の鮮度が高く、かつ、ユーザのニーズに適した検索結果を提供することができるという効果を奏する。

【0088】請求項8の発明に係る情報源観測方法は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測方法において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存するステップと、上記参照情報を、参照元となる各URL毎に集計するステップとを有している。

【0089】これにより、参照元として頻繁に利用されるURLを把握することが可能となるという効果を奏する。すなわち、ユーザにとって有用なURLを的確に提示することが可能となるという効果を奏する。

【0090】請求項9の発明に係る情報源観測処理を実行させるプログラムを記録した記録媒体は、ネットワークで繋がれたサーバコンピュータ上のハイパーリンク構造を有するデータファイルをクライアントコンピュータ上で閲覧する際の情報源観測処理を実行させるプログラムを記録した記録媒体において、クライアントコンピュータからアクセスがあったデータファイルの参照元のURLを参照情報として保存する処理と、上記参照情報を、参照元となる各URL毎に集計する処理とをコンピュータに実行させるためのプログラムを記録した、コンピュータ読み取り可能である構成である。

【0091】これにより、参照元として頻繁に利用されるURLを把握することが可能なプログラムを提供することができるという効果を奏する。すなわち、ユーザに

にとって有用なURLを的確に提示することが可能なプログラムを提供することができるという効果を奏する。

【図面の簡単な説明】

【図1】本発明の実施の一形態に係る情報源観測手段の概略構成、および、該情報源観測手段と、それに接続されるサーバコンピュータおよびクライアントコンピュータとの関係を示すブロック図である。

【図2】上記情報源観測手段がクライアントコンピュータに内蔵された場合の概略構成を示すブロック図である。

【図3】上記情報源観測手段の構成要素の概略を示すブロック図である。

【図4】クライアントコンピュータから発行される要求ヘッダの一例を示す説明図である。

【図5】サーバコンピュータから発行される応答ヘッダの一例を示す説明図である。

【図6】情報源観測手段における処理の流れを示すフローチャートである。

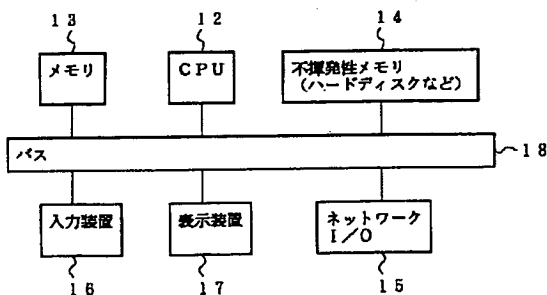
【図7】情報源観測手段が備えるデータ変換手段によって生成されたデータを表示させたときの一例を示す説明図である。

【図8】WWW上におけるHTMLページおよびデータファイルのリンクの様子の一例を示す概念図である。

【符号の説明】

- 1 サーバコンピュータ
- 2 情報源観測手段（情報源観測装置）
- 3 クライアントコンピュータ
- 4 参照情報抽出手段
- 5 参照情報保存手段
- 6 カウント手段
- 7 ソート手段
- 8 データ変換手段
- 9 情報更新手段
- 10 情報蓄積手段
- 11 検索手段

【図3】



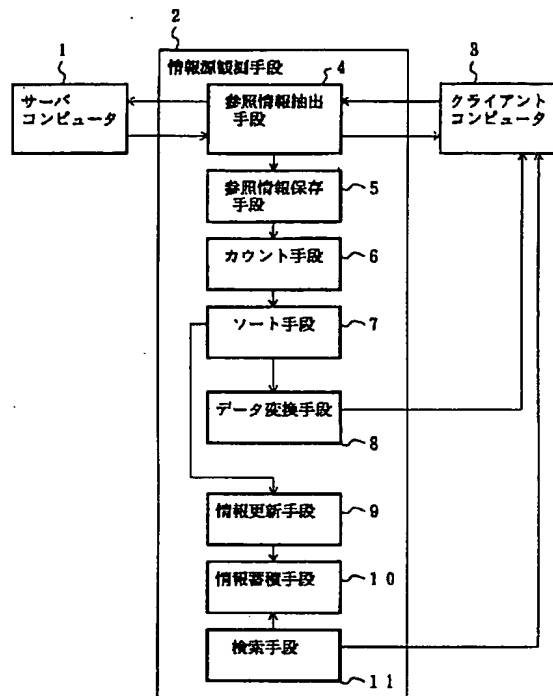
【図4】

```
GET http://www.news/1.html HTTP/1.0
Referer: http://www.news/
```

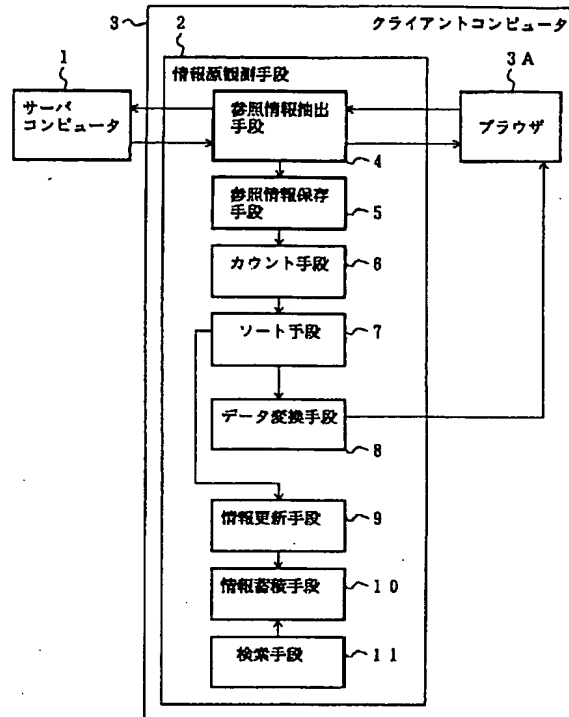
【図5】

```
HTTP/1.0 200 Document Follows
Date: Fri, 23 Oct 1998 004:23:34 GMT
Content-Type: text/html
(空行)
テキストデータ (HTMLデータ) 本体
```

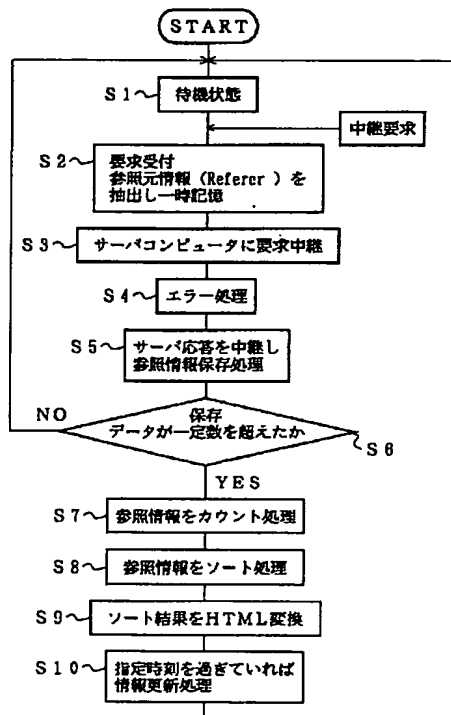
【図1】



【図2】



【図6】



【図7】

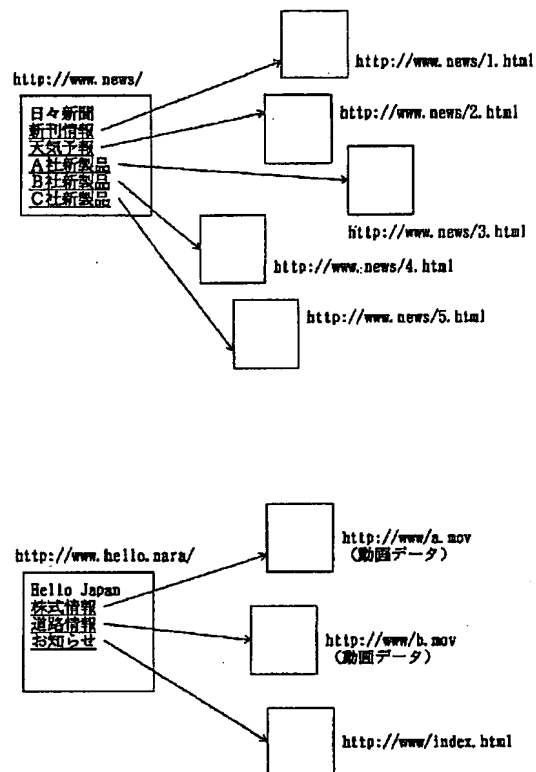
あなたの情報源リスト

1998年10月27日18:30 更新

順位	URL	タイトル
1	http://www.news/	日々新聞
2	http://www.hello.nara/	奈良県地域情報
3	http://www.cnn.com/	CNN Interactive
4	http://www.asahi.com/	朝日新聞ニュース

順位	URL	タイトル
1	http://www.hello.nara/	奈良県地域情報
2	http://www.movie.archive/	Movie archive

【図8】



フロントページの続き

(51)Int.Cl.⁷

識別記号

F I
G 0 6 F 15/419

テーマコード(参考)
3 2 0